A Truthful Incentive Scheme for Federated Learning with Strategic Rejection

Mingshu Cong[1,3*†] , **Zhongming Qu**[2,3*] , **Han Yu**[4] , **Xi Weng**[5] , **Zichen Chen**[4] , **Jiabao Qu**[3] , **Siu Ming Yiu**[1] , **Yang Liu**[6] and

[1]The FinTech and Blockchain Laboratory, The University of Hong Kong, Hong Kong
[2]Hong Kong University of Science and Technology, Hong Kong
[3]LogiOcean, Shenzhen, China
[4]Nanyang Technological University (NTU), Singapore
[5]Peking University, Beijing, China
[6]WeBank, Shenzhen, China
[*]Equal contribution [†]Corresponding author
miranda.cong@logiocean.com, zhongming@logiocean.com, han.yu@ntu.edu.sg,
wengxi125@gsm.pku.edu.cn, zichen002@e.ntu.edu.sg, qujiabao@logiocean.com smyiu@cs.hku.hk
yangliu@webank.com

Abstract

Federated learning (FL) enables privacy-preserving collaborative model training. When businesses with potentially competing interests join FL, it is crucial to compensate them for their economic costs incurred in order to sustain participation. At the same time, the FL coordinator needs to be selective in order to retain participants who contribute positively to the federation. Two types of information asymmetry in FL hinders these objectives: 1) the potential usefulness of participants' original datasets and 2) private costs incurred by FL participants. In this paper, we propose a Vickrey–Clarke–Groves-based FL incentive mechanism, FVCG, to address these problems by recommending the FL coordinator to selectively reject costly data based on ex-ante evaluations of the usefulness of the offered data. Such a right-of-rejection encourages data owners to offer their original high-quality datasets and truthfully report their cost types. FVCG provides theoretical guarantees for incentive compatibility and social surplus maximization. It minimizes expected deviations from individual rationality, is by expectation weakly budget balanced, and supports user-defined fairness criteria.

1 Introduction

Data is key to building artificial intelligence (AI) empowered applications. However, in practice, data are collected and stored by different entities which are either unwilling or unable to share data directly with others [Yang *et al.*, 2019a; Kairouz *et al.*, 2019; Yang *et al.*, 2019b]. Federated learning (FL) has emerged in recent years as an alternative solution to build AI models based on distributedly stored data while preserving data privacy.

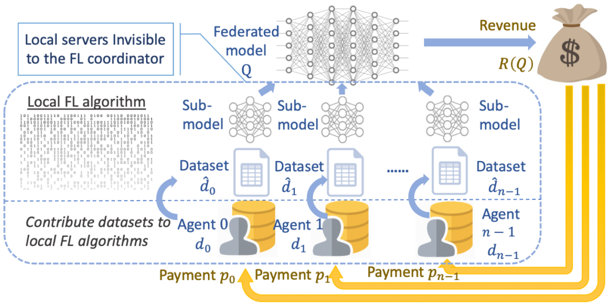

Figure 1: The flow of incentives in federated learning

Nevertheless, FL may not always be profitable for participants. This is especially the case for businesses participating in FL. Besides the cost of gathering and cleaning data, business participants also bear technical, compliance and opportunity costs (e.g., due to their market share erosion) [Yu *et al.*, 2020]. In addition, there exists the problem of free riding. If every participant enjoys the benefit of the FL model equally, a rational participant may only join FL training with its low-quality data and conceal high-quality data so as to protect its competitive advantage. Without a properly designed incentive scheme, the aforementioned problems will threaten the sustainability of a federation over time.

In FL, model parameters are updated based on locally stored datasets from participants (i.e., agents) in an encrypted manner. Although no raw data are being physically transmitted from the participants to the federation, for simplicity of terms, we refer to "joining federated model training" and "contributing data" interchangeably in this paper. We propose the design of an incentive scheme as illustrated by Figure 1 to incentivize participants. As shown by the figure, revenue generated by the FL model is distributed among participants to cover their costs of contributing data. Here, there is asymmetric information about both the truthfulness of the contributed dataset and the cost. Similar problems in data collaboration have been discussed by [Faltings and Radanovic, 2017]. An FL incentive scheme needs to incentivize participants to offer their original high-quality datasets and truthfully report their private cost types.

In order to promote truthfulness of participants in FL model training, we propose an incentive scheme inspired by *buyer's right of rejection*. A buyer has the right to reject defective goods. Such an approach transfers the burden of ensuring the quality of goods from the buyer to the seller. In FL, goods are data involved in FL model training; the sellers are the participants; the buyer is the FL coordinator.

In particular, we propose FVCG, a Federated VCG (Vickrey–Clarke–Groves as proposed by [Vickrey, 1961; Clarke, 1971; Groves, 1973]) incentive scheme, which entitles the FL coordinator with the right to reject high-cost-low-quality data. Under FVCG, the FL coordinator firstly evaluates the usefulness of a participant's offered data in a sandbox and analyzes it against its reported cost to determine if strategic rejection shall be triggered. Once the participant passes this round of evaluation, FVCG computes the payoff for this participant, which consists of the VCG payment component (which guarantees incentive compatibility and social surplus maximization) and an adjustment payment component (which promotes individual rationality, weak budget balancedness, and accommodates user-defined fairness criteria). As the adjustment payment may not be in a closed form, we propose a neural-network-based learning algorithm to compute its value.

Theoretical analysis proves that, under some reasonable conditions, FVCG simultaneously maximizes social surplus, minimizes expected deviations from individual rationality, is by expectation weakly budget balanced, and supports user-defined fairness. Extensive experimental evaluation shows that compared to state-of-the-art approaches, FVCG effectively incentivizes participants to offer their original datasets, keeps all participants in the federation, results in higher social surplus, and brings profits to the FL coordinator.

2 Related Works

In recent years, there has been a growing body of literature on training machine learning models with strategic participants. This line of research can be divided into three groups.

The first group assumes that predictions from the trained model affect participants' utilities, and hence they may report false data to drive model predictions closer to their expectations [Perote and Perote-Pena, 2004; Dekel *et al.*, 2010; Caragiannis *et al.*, 2016; Meir *et al.*, 2012]. The second group focuses on the loss of privacy and compensates data providers for such privacy losses [Ghosh and Roth, 2015; Cummings *et al.*, 2015; Nissim *et al.*, 2012; Fleischer and Lyu, 2012]. The third group views data as economic resources with both values and costs. The model builder pays monetary incentives to encourage data providers to contribute higher-quality data with higher values. The value of a dataset can be measured by its influence on the performance of the trained model (e.g., its Shapley value) [Jia *et al.*, 2019; Wang, 2019; Cai *et al.*, 2015], with considerations on the costs incurred [Richardson *et al.*, 2019; Westenbroek *et al.*, 2019; Yu *et al.*, 2020].

FVCG belongs to the third group. Compared to previous works, FVCG is unique as it focuses on FL scenarios involving business participants which are assumed to be more strategic and thus, presents a more complex game-theoretic environment. Methodologies in these papers are adopted from *Mechanism Design*, an important subfield of game theory [Nisan *et al.*, 2007; Koutsoupias, 2014]. FVCG is based on the famous VCG mechanism [Vickrey, 1961; Clarke, 1971; Groves, 1973] and the literature on the computational aspect of VCG mechanisms [Nisan and Ronen, 2007]. However, unlike the classical VCG mechanism, FVCG can achieve individual rationality and weak budget balancedness for rich type spaces under certain conditions, releasing the tension between these objectives explained in [Jackson, 2014].

3 Preliminaries

We consider a problem with a set of n *participants* denoted by $N = \{0, 1, \ldots, n-1\}$. Each participant $i \in N$ owns a private dataset. Each participant contributes to the federated model quality Q with parameter-updates based on the whole or part of his original dataset. The federated model can generate revenue $R(Q)$, which is to be distributed to the participants. Each participant receives a payment p_i.

A dataset is measured by its *usefulness*, which represents how useful the dataset is in updating a given FL model. The usefulness of the original private dataset and the contributed dataset are denoted by d_i and $\hat{d}_i$, respectively. d_i and $\hat{d}_i$ can be vectors composing of multiple measures of usefulness. Since a participant may contribute only a subset of the original dataset, we have $\hat{d}_i \leq d_i$, where $\geq$ denotes the element-wise vector comparison. We also set the usefulness of an empty dataset to be $\mathbf{0}$, and the original dataset satisfies $d_i > \mathbf{0}$. The federated model quality Q is assumed to be a function of $\hat{\boldsymbol{d}} = (\hat{d}_0, \ldots, \hat{d}_{n-1})$. The federation revenue R, which depends on Q, also depends on $\hat{\boldsymbol{d}}$. We use $B(\cdot)$ to denote the composite function, i.e., $B(\hat{\boldsymbol{d}}) = R(Q(\hat{\boldsymbol{d}}))$.

Joining FL model training incurs costs to participants. It is reasonable to assume i's cost c_i to increase with $\hat{d}_i$. Since the cost of contributing a dataset with the same usefulness may vary among participants, we introduce another parameter γ_i, called *cost type* such that $c_i = c_i(\hat{d}_i, \gamma_i)$. In our mechanism, participants report their private cost types to the FL coordinator when offering their datasets. The reported cost type of i is denoted by $\hat{\gamma}_i$. The participant's preference is represented by the *quasi-linear utility* $u_i = p_i - c_i$. Since we focus on dominant strategy mechanism design, there is no need to impose any assumptions on either the prior distributions of participants' types (d_i, γ_i), or participants' knowledge of each other's type distribution. Nevertheless, lacking real behavioral data of participants in practice, we may require such prior distributions to generate synthetic data for the training of neural networks.

A federation desires to maximize the *social surplus* $S(\hat{\boldsymbol{d}}, \boldsymbol{\gamma}) = B(\hat{\boldsymbol{d}}) - \sum_{i=0}^{n-1} c_i(\hat{d}_i, \gamma_i)$, which is a classical economic measure defined as the total benefit minus the total cost in an economic system. Social surplus maximization implies

Pareto efficiency. As the social surplus depends on the usefulness of contributed datasets, the FL coordinator cannot optimize it without the right-of-rejection. In FVCG, the FL coordinator can reject a participant fully or partially in a given round of FL model training. The usefulness of the offered dataset by participant i is denoted by $\check{d}_i$. The usefulness of the accepted/contributed dataset is no greater than the usefulness of the offered dataset (i.e., $\hat{d}_i \leq \check{d}_i$). These two usefulness measures are related by $\eta_i = \hat{d}_i \oslash \check{d}_i \in [0,1]^{dim(d_i)}$ and $\hat{d}_i = \check{d}_i \odot \eta_i$, where $\odot$ and $\oslash$ denote the element-wise multiplication and division, respectively. η_i is the *acceptance level*, which is a continuous value in the $dim(d_i)$-dimensional cube $[0,1]^{dim(d_i)}$. The FL coordinator controls the acceptance level. For example, $\eta_i = \mathbf{0}$ means that the offered dataset is entirely rejected, whereas $\eta_i = \mathbf{1}$ means that the offered dataset is entirely accepted. $\mathbf{0} < \eta_i < \mathbf{1}$ means the offered dataset is partially accepted.

In order for the FL coordinator to evaluate the usefulness of a participant's offered dataset before accepting it, we assume the existence of a *sandbox*. A sandbox is a testing environment isolated from the production environment. A copy of the current version of the FL model can be placed in the sandbox, based on which participants can evaluate the usefulness of their datasets through algorithms such as [Koh and Liang, 2017]. As the design of the sandbox is not the focus of this paper, we assume its availability and FVCG can obtain the usefulness of a given participant's offered dataset from it.

4 The Proposed FVCG Approach

The FVCG payment p_i to participant i is composed of the *VCG payment* τ_i and the *adjustment payment* $h_i^* + g_i^*$. FVCG computes these payments through the following procedure.

4.1 Computing the Optimal Acceptance Levels

To determine the optimal level of acceptance for each participant's local dataset, we optimize the following objective function:

$$\begin{aligned}\boldsymbol{\eta}^* &= \mathrm{argmax}_{\boldsymbol{\eta} \in [0,1]^{\mathrm{dim}(d_i) \times n}} \{S(\check{\boldsymbol{d}} \odot \boldsymbol{\eta}, \hat{\boldsymbol{\gamma}})\} \\ &= \mathrm{argmax}_{\boldsymbol{\eta} \in [0,1]^{\mathrm{dim}(d_i) \times n}} \{B(\check{\boldsymbol{d}} \odot \boldsymbol{\eta}) - \sum_{i=0}^{n-1} c_i(\check{d}_i \odot \eta_i, \hat{\gamma}_i)\},\end{aligned} \tag{1}$$

where $\boldsymbol{\eta}^* = (\eta_0^*, \dots, \eta_{n-1}^*)$. Different $(\check{\boldsymbol{d}}, \hat{\boldsymbol{\gamma}})$ results in different $\boldsymbol{\eta}^*$. Hence, $\boldsymbol{\eta}^*$ is written as $\boldsymbol{\eta}^*(\check{\boldsymbol{d}}, \hat{\boldsymbol{\gamma}})$.

The maximum social surplus is denoted by $S^*(\check{\boldsymbol{d}}, \hat{\boldsymbol{\gamma}}) = B(\check{\boldsymbol{d}} \odot \boldsymbol{\eta}^*(\check{\boldsymbol{d}}, \hat{\boldsymbol{\gamma}})) - \sum_{i=0}^{n-1} c_i(\check{d}_i \odot \eta_i^*(\check{\boldsymbol{d}}, \hat{\boldsymbol{\gamma}}), \hat{\gamma}_i)$. Although $S^*(\check{\boldsymbol{d}}, \hat{\boldsymbol{\gamma}})$ and $S(\hat{\boldsymbol{d}}, \boldsymbol{\gamma})$ both represent social surplus, they are different functions. The independent variables of $S(\cdot)$ are the usefulness of the contributed dataset $\hat{\boldsymbol{d}}$ and the true cost types $\boldsymbol{\gamma}$, whereas the independent variables of $S^*(\cdot)$ are the usefulness of the offered dataset $\check{\boldsymbol{d}}$ and the reported cost types $\hat{\boldsymbol{\gamma}}$.

4.2 Computing the VCG Payments

With $\boldsymbol{\eta}^*$, the VCG payment τ_i to participant i can be computed as:

$$\begin{aligned}\tau_i &= S^*(\check{\boldsymbol{d}}, \hat{\gamma}) - S_{-i}^*(\check{\boldsymbol{d}}_{-i}, \hat{\boldsymbol{\gamma}}_{-i}) + c_i(\check{d}_i \odot \eta_i^*(\check{\boldsymbol{d}}, \hat{\boldsymbol{\gamma}}), \hat{\gamma}_i) \\ &= [B(\check{\boldsymbol{d}} \odot \boldsymbol{\eta}^*(\check{\boldsymbol{d}}, \hat{\boldsymbol{\gamma}})) - B(\check{\boldsymbol{d}}_{-i} \odot \boldsymbol{\eta}^{-i*}(\check{\boldsymbol{d}}_{-i}, \hat{\boldsymbol{\gamma}}_{-i}))] \\ &\quad - \sum_{k \neq i}[c_k(\check{d}_k \odot \eta_k^*(\check{\boldsymbol{d}}, \hat{\boldsymbol{\gamma}}), \hat{\gamma}_k) - c_k(\check{d}_k \odot \eta_k^{-i*}(\check{\boldsymbol{d}}_{-i}, \hat{\boldsymbol{\gamma}}_{-i}), \hat{\gamma}_k)],\end{aligned} \tag{2}$$

where $(\check{\boldsymbol{d}}_{-i}, \hat{\boldsymbol{\gamma}}_{-i})$ denotes the offered data usefulness and reported cost types excluding participant i, and $\boldsymbol{\eta}^{-i*}$ and $S_{-i}^*(\check{\boldsymbol{d}}_{-i}, \hat{\boldsymbol{\gamma}}_{-i})$ are the corresponding optimal data acceptance levels and social surplus. $\boldsymbol{\tau}$ is a function of $(\check{\boldsymbol{d}}, \hat{\boldsymbol{\gamma}})$, written as $\boldsymbol{\tau}(\check{\boldsymbol{d}}, \hat{\boldsymbol{\gamma}})$. Note that $\boldsymbol{\eta}^{-i*}$ is different from $\boldsymbol{\eta}_{-i}^*$. The former maximizes $S(\check{\boldsymbol{d}}_{-i} \odot \boldsymbol{\eta}_{-i}, \hat{\boldsymbol{\gamma}}_{-i})$, whereas the latter is the component of $\boldsymbol{\eta}^*$ that maximizes $S(\check{\boldsymbol{d}} \odot \boldsymbol{\eta}, \hat{\boldsymbol{\gamma}})$.

4.3 Learning the Optimal Adjustment Payments

The adjustment payments are introduced to promote individual rationality, weak budget balancedness, and user-defined fairness criteria, without hurting incentive compatibility and social surplus maximization guaranteed by the VCG payments. We set the adjustment payment to participant i as:

$$h_i(\check{\boldsymbol{d}}_{-i}, \hat{\boldsymbol{\gamma}}_{-i}) + g_i(\check{d}_i), \tag{3}$$

where $h_i(\cdot)$ is a function of $(\check{\boldsymbol{d}}_{-i}, \hat{\boldsymbol{\gamma}}_{-i})$, and $g_i(\cdot)$ is an increasing function of $\check{d}_i$.

As in the standard VCG mechanism, adding the adjustment payment $h_i(\cdot)$ in Eq. 3 to the VCG payment does not affect incentive compatibility and social surplus maximization. But different from the standard VCG mechanism, we also add $g_i(\cdot)$, which may affect incentive compatibility. So we require $g_i(\cdot)$ to be increasing to guarantee that offering the original dataset with usefulness d_i is incentive compatible. The optimal adjustment payment functions $h_i^*(\cdot)$ and $g_i^*(\cdot)$ are such chosen that the total payment to participant i, calculated by

$$p_i(\cdot) = \tau_i(\cdot) + h_i^*(\cdot) + g_i^*(\cdot), \tag{4}$$

minimizes the deviations from individual rationality (i.e., Loss1), weak budget balancedness (i.e., Loss2), and user defined objectives such as fairness (i.e., Loss3) in the following loss function:

$$\mathrm{LOSS} = \lambda_1 \mathrm{Loss1} + \lambda_2 \mathrm{Loss2} + \lambda_3 \mathrm{Loss3}. \tag{5}$$

The expressions of Loss1 and Loss2 will be provided in the next section. Loss3 can be any reasonable user-given unfairness function. For instance:

$$\mathrm{Loss3} = \frac{1}{n} \sum_{i=0}^{n-1} \left(\frac{\check{d}_i}{\check{d}_i + p_i} \right)^2 - \left[\frac{1}{n} \sum_{i=0}^{n-1} \left(\frac{\check{d}_i}{\check{d}_i + p_i} \right) \right]^2. \tag{6}$$

Such a Loss3 is free from the division-by-zero problem and drives the unit price for offered data usefulness to be the same for all participants. Nevertheless, the specific form of Loss3 has no effect on the theoretical guarantees for objectives other than this user-defined fairness.

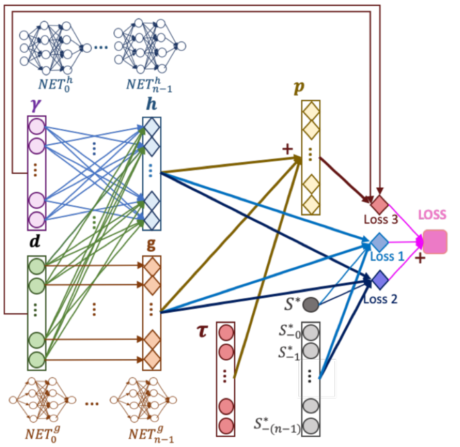

Figure 2: The composite neural network of FVCG

Because there is no closed-form solution for the optimal $h_i^*(\cdot)$ and $g_i^*(\cdot)$, we approximate them with neural networks. In particular, we construct n neural networks $\mathrm{NET}_i^h, i \in N$, which share the same set of parameters, to approximate $h_i(\cdot), i \in N$, and another n monotonic networks (i.e., neural networks with all weights being non-negative) $\mathrm{NET}_i^g, i \in N$, sharing another set of parameters, to approximate $g_i(\cdot), i \in N$. Output nodes of these $2n$ networks, denoted by $\mathrm{NET}_i^h.o, \mathrm{NET}_i^g.o, i \in N$, are combined into a single *composite neural network* in Figure 2 with the following loss function (of which the reasonableness will be proved in the next section):

$$\begin{aligned}\mathrm{LOSS} = \frac{1}{T}\sum_{t=0}^{T-1}\{\lambda_1 \sum_{i=0}^{n-1} \mathrm{ReLu}[-(S^{*t} - S^{*t}_{-i})\\ - (\mathrm{NET}_i^h.o + \mathrm{NET}_i^g.o)] + \lambda_2 \mathrm{ReLu}[\sum_{i=0}^{n-1}(\qquad (7)\\ [(S^{*t} - S^{*t}_{-i}) + (\mathrm{NET}_i^h.o + \mathrm{NET}_i^g.o)] - S^{*t})]\\ + \lambda_3 \mathrm{Loss3}(\boldsymbol{\tau}^t + \mathrm{NET}_i^h.o + \mathrm{NET}_i^g.o, \boldsymbol{d}^t)\},\end{aligned}$$

where T is the sample size. For the sample t, $\boldsymbol{\tau}^t = \boldsymbol{\tau}(\boldsymbol{d}^t, \boldsymbol{\gamma}^t)$, $S^{*t}_{-i} = S^*(\boldsymbol{d}^t_{-i}, \boldsymbol{\gamma}^t_{-i})$, and $S^{*t} = S^*(\boldsymbol{d}^t, \boldsymbol{\gamma}^t)$.

The composite neural network can be trained based on participants' behaviors in real scenarios. If such behavioral data are not available, FL coordinators may leverage domain experts' knowledge to generate synthetic data $(\boldsymbol{d}^t, \boldsymbol{\gamma}^t)$ from the estimated prior distribution $(\Delta(\boldsymbol{d}), \Delta(\boldsymbol{\gamma}))$ during the bootstrapping stage.

The training procedure for FVCG is described in Algorithm 1. Within this algorithm, CAPITALIZED words refer to nodes in the composite neural network, whereas *lowercased* words stand for other types of variables. The value of a node is referenced as NODE.$value$. The composite neural network is constructed according to Eq. 7, where values of the following variables are fed to the input nodes:

$$(\boldsymbol{d}^t, \boldsymbol{\gamma}^t), \boldsymbol{\tau}^t, S^{*t}_{-i}, S^{*t}, i \in N, t = 0, \ldots, T-1. \qquad (8)$$

Algorithm 1 Train $[\mathrm{NET}_i^h]$, $[\mathrm{NET}_i^g]$ for all $i \in N$

```
Require: Neural networks: $[\mathrm{NET}_i^h]$, $[\mathrm{NET}_i^g]$
    Probability Distribution: $\Delta(\boldsymbol{d})$ , $\Delta(\boldsymbol{\gamma})$
Ensure: Trained networks: $[\mathrm{NET}_i^h]$, $[\mathrm{NET}_i^g]$
construct the composite neural network for LOSS where input nodes are those in Eq. 8 and the LOSS node is constructed according to Eq. 7
PAR ← list of all parameters of the computational graph
initialize PAR
for $l$ in range($iterations$) do
    for $t$ in range($T$) do
        randomly draw $(\boldsymbol{d}^t, \boldsymbol{\gamma}^t)$ from $\Delta(\boldsymbol{d}), \Delta(\boldsymbol{\gamma})$
        compute $\boldsymbol{\tau}^t$, $S^{*t}$ and all $S^{*t}_{-i}$
    end for
    assign $(\boldsymbol{d}^t, \boldsymbol{\gamma}^t), \boldsymbol{\tau}^t, S^{*t}_{-i}, S^{*t}, i \in N, t = 1, \ldots, S$ to corresponding input nodes of LOSS.
    feedForward(LOSS)
    backPropogation(LOSS)
    $\delta \leftarrow (\frac{d\mathrm{LOSS}}{d\mathrm{PAR}}).value$
    PAR.$value$ ← PAR.$value - a * \delta$
    if $\tau_i^t + \mathrm{NET}_i^{h,t}.o.value + \mathrm{NET}_i^{g,t}.o.value < 0$ then
        increase the bias of the output layer of $\mathrm{NET}_i^h$ by $b$
    end if
end for
```

Finally, the FVCG payment to participant i is the sum of the VCG payment and the optimal adjustment payment:

$$p_i = \tau_i(\check{\boldsymbol{d}}, \hat{\boldsymbol{\gamma}}) + \mathrm{NET}_i^h.o + \mathrm{NET}_i^g.o, \qquad (9)$$

where $(\check{\boldsymbol{d}}_{-i}, \hat{\boldsymbol{\gamma}}_{-i})$ and $\check{d}_i$ are fed into the input nodes of the trained networks NET_i^h and NET_i^g, respectively.

5 Analytical Evaluation

In this section, we analyze the properties of FVCG. We first prove for arbitrary $h_i(\check{\boldsymbol{d}}_{-i}, \hat{\boldsymbol{\gamma}}_{-i})$ and increasing $g_i(\hat{d}_i)$, the payment $p_i(\cdot) = \tau_i(\cdot) + h_i(\cdot) + g_i(\cdot)$ satisfies dominant strategy incentive compatibility and social surplus maximization.

Proposition 1 (Dominant strategy incentive compatibility)**.** *For every participant i, truthfully reporting its cost type γ_i and offering its original dataset for FL model training is the dominant strategy under FVCG, i.e.,*

$$\begin{aligned}p_i((d_i, \check{\boldsymbol{d}}_{-i}), (\gamma_i, \hat{\boldsymbol{\gamma}}_{-i})) - c_i(d_i \odot \eta_i^*((d_i, \check{\boldsymbol{d}}_{-i}), (\gamma_i, \hat{\boldsymbol{\gamma}}_{-i})), \gamma_i)\\ \geq p_i(\check{\boldsymbol{d}}, \hat{\boldsymbol{\gamma}}) - c_i(\check{d}_i \odot \eta_i^*(\check{\boldsymbol{d}}, \hat{\boldsymbol{\gamma}}), \gamma_i), \forall d_i, \check{\boldsymbol{d}}, \gamma_i, \hat{\boldsymbol{\gamma}}. \qquad (10)\end{aligned}$$

Proof. By definition of $S^*(\check{\boldsymbol{d}}, \hat{\boldsymbol{\gamma}})$ (i.e., $S^*(\check{\boldsymbol{d}}, \hat{\boldsymbol{\gamma}}) = \max_{\boldsymbol{\eta} \in [0,1]^{\dim(d_i) \times n}}\{B(\check{\boldsymbol{d}} \odot \boldsymbol{\eta}) - \sum_{i=0}^{n-1} c_i(\check{d}_i \odot \eta_i, \hat{\gamma}_i)\}, \forall \check{\boldsymbol{d}}, \hat{\boldsymbol{\gamma}}$), we substitute $\check{\boldsymbol{d}}$ with $(d_i, \check{\boldsymbol{d}}_{-i})$ and get

$$\begin{aligned}S^*((d_i, \check{\boldsymbol{d}}_{-i}), (\gamma_i, \hat{\boldsymbol{\gamma}}_{-i})) \geq B((d_i, \check{\boldsymbol{d}}_{-i}) \odot \boldsymbol{\eta})\\ - c_i(d_i \odot \eta_i, \gamma_i) - \sum_{k \neq i} c_i(\check{d}_k \odot \eta_k, \hat{\gamma}_k), \forall \boldsymbol{\eta}. \qquad (11)\end{aligned}$$

In particular, for $\boldsymbol{\eta} = (\check{d}_i \odot \eta_i^*(\check{\boldsymbol{d}}, \hat{\boldsymbol{\gamma}}) \oslash d_i, \boldsymbol{\eta}^*_{-i}(\check{\boldsymbol{d}}, \hat{\boldsymbol{\gamma}})) \in [0,1]^{\dim(d_i) \times n}$, Eq. 11 holds. Also, because the usefulness of the offered dataset (which is verified in the sandbox) is no greater than the original dataset, $d_i \geq \check{d}_i$ and $g(d_i) \geq g(\check{d}_i)$. Therefore,

$$\begin{aligned}
&S^*((d_i, \check{\boldsymbol{d}}_{-i}), (\gamma_i, \hat{\boldsymbol{\gamma}}_{-i})) + g_i(d_i) \\
\geq &B((d_i, \check{\boldsymbol{d}}_{-i}) \odot (\check{d}_i \odot \eta_i^*(\check{\boldsymbol{d}}, \hat{\boldsymbol{\gamma}}) \oslash d_i, \boldsymbol{\eta}_{-i}^*(\check{\boldsymbol{d}}, \hat{\boldsymbol{\gamma}}))) \\
&- \sum_{k \neq i} c_i(\check{d}_k \odot \eta_k^*(\check{\boldsymbol{d}}, \hat{\boldsymbol{\gamma}}), \hat{\gamma}_k) + g_i(\check{d}_i) \\
&- c_i(d_i \odot \check{d}_i \odot \eta_i^*(\check{\boldsymbol{d}}, \hat{\boldsymbol{\gamma}}) \oslash d_i, \gamma_i) \qquad (12) \\
= &S((\check{d}_i, \check{\boldsymbol{d}}_{-i}) \odot \boldsymbol{\eta}^*(\check{\boldsymbol{d}}, \hat{\boldsymbol{\gamma}}), (\hat{\gamma}_i, \hat{\boldsymbol{\gamma}}_{-i})) + g_i(\check{d}_i) \\
&+ c_i(\check{d}_i \odot \eta_i^*(\check{\boldsymbol{d}}, \hat{\boldsymbol{\gamma}}), \hat{\gamma}_i) - c_i(\check{d}_i \odot \eta_i^*(\check{\boldsymbol{d}}, \hat{\boldsymbol{\gamma}}), \gamma_i) \\
= &S^*(\check{\boldsymbol{d}}, \hat{\boldsymbol{\gamma}}) + g_i(\check{d}_i) \\
&+ c_i(\check{d}_i \odot \eta_i^*(\check{\boldsymbol{d}}, \hat{\boldsymbol{\gamma}}), \hat{\gamma}_i) - c_i(\check{d}_i \odot \eta_i^*(\check{\boldsymbol{d}}, \hat{\boldsymbol{\gamma}}), \gamma_i).
\end{aligned}$$

Adding $h_i(\check{\boldsymbol{d}}_{-i}, \hat{\boldsymbol{\gamma}}_{-i}) - S_{-i}^*(\check{\boldsymbol{d}}_{-i}, \hat{\boldsymbol{\gamma}}_{-i})$ to both sides of Eq. 12 and substituting $p_i(\cdot) = S^*(\cdot) - S_{-i}^*(\cdot) + c_i(\cdot) + h(\cdot) + g(\cdot)$, we get

$$\begin{aligned}
&p_i((d_i, \check{\boldsymbol{d}}_{-i}), (\gamma_i, \hat{\boldsymbol{\gamma}}_{-i})) - c_i(d_i \odot \eta_i^*((d_i, \check{\boldsymbol{d}}_{-i}), (\gamma_i, \hat{\boldsymbol{\gamma}}_{-i})), \gamma_i) \\
&\geq p_i(\check{\boldsymbol{d}}, \hat{\boldsymbol{\gamma}}) - c_i(\check{d}_i \odot \eta_i^*(\check{\boldsymbol{d}}, \hat{\boldsymbol{\gamma}}), \gamma_i). \qquad (13)
\end{aligned}$$

□

Proposition 2 (Social surplus maximization)**.** *FVCG maximizes the social surplus.*

Proof. Suppose $\hat{\boldsymbol{d}}^{**} = \text{argmax}_{\hat{\boldsymbol{d}} \leq \boldsymbol{d}} \{S(\hat{\boldsymbol{d}}, \boldsymbol{\gamma})\}$ and $S^{**} = S(\hat{\boldsymbol{d}}^{**}, \boldsymbol{\gamma})$. We aim to prove that FVCG results in social surplus no less than S^{**}.

By definition of $\boldsymbol{\eta}^*(\check{\boldsymbol{d}}, \hat{\boldsymbol{\gamma}})$,

$$S(\check{\boldsymbol{d}} \odot \boldsymbol{\eta}^*(\check{\boldsymbol{d}}, \hat{\boldsymbol{\gamma}}), \hat{\boldsymbol{\gamma}}) \geq S(\check{\boldsymbol{d}} \odot \boldsymbol{\eta}, \hat{\boldsymbol{\gamma}}), \forall \boldsymbol{\eta}. \qquad (14)$$

Incentive compatibility means $\check{\boldsymbol{d}} = \boldsymbol{d}$, $\hat{\boldsymbol{\gamma}} = \boldsymbol{\gamma}$. Therefore,

$$S(\boldsymbol{d} \odot \boldsymbol{\eta}^*(\boldsymbol{d}, \boldsymbol{\gamma}), \boldsymbol{\gamma}) \geq S(\boldsymbol{d} \odot \boldsymbol{\eta}, \boldsymbol{\gamma}), \forall \boldsymbol{\eta}. \qquad (15)$$

Particularly, Eq. 15 holds for $\boldsymbol{\eta} = \hat{\boldsymbol{d}}^{**} \oslash \boldsymbol{d} \in [0,1]^{\dim(d_i) \times n}$, i.e.,

$$\begin{aligned}
S(\boldsymbol{d} \odot \boldsymbol{\eta}^*(\boldsymbol{d}, \hat{\boldsymbol{\gamma}}), \boldsymbol{\gamma}) &\geq S(\boldsymbol{d} \odot \hat{\boldsymbol{d}}^{**} \oslash \boldsymbol{d}, \boldsymbol{\gamma}) \\
&= S(\hat{\boldsymbol{d}}^{**}, \boldsymbol{\gamma}) = S^{**}. \qquad (16)
\end{aligned}$$

The left side of Eq. 16 is the social surplus achieved by FVCG, while the right side is the maximum possible social surplus for the given $(\boldsymbol{d}, \boldsymbol{\gamma})$. Therefore, the left side equals the right side. □

Based on the dominant incentive compatibility guaranteed by Proposition 1, we can give conditions for individual rationality and weak budget balancedness.

Proposition 3 (Individual rationality)**.** *FVCG is individual rational (IR) i.f.f. the usefulness of the original datasets and the true cost types satisfy*

$$h_i(\boldsymbol{d}_{-i}, \boldsymbol{\gamma}_{-i}) + g_i(d_i) \geq -[S^*(\boldsymbol{d}, \boldsymbol{\gamma}) - S_{-i}^*(\boldsymbol{d}_{-i}, \boldsymbol{\gamma}_{-i})]. \qquad (17)$$

Proof. Since playing truthfully is shown to be the dominant strategy by Proposition 1, we use $\boldsymbol{d}, \boldsymbol{\gamma}$ to substitute $\check{\boldsymbol{d}}, \hat{\boldsymbol{\gamma}}$ in Eq. 2 and Eq. 4. Then, the utility of participant i becomes

$$\begin{aligned}
u_i(\boldsymbol{d}, \boldsymbol{\gamma}) &= p_i(\boldsymbol{d}, \boldsymbol{\gamma}) - c_i(d_i \odot \eta_i^*(\boldsymbol{d}, \boldsymbol{\gamma}), \gamma_i) \qquad (18) \\
&= S^*(\boldsymbol{d}, \boldsymbol{\gamma}) - S_{-i}^*(\boldsymbol{d}_{-i}, \boldsymbol{\gamma}_{-i}) + h_i(\boldsymbol{d}_{-i}, \boldsymbol{\gamma}_{-i}) + g_i(d_i).
\end{aligned}$$

IR requires $u_i(\boldsymbol{d}, \boldsymbol{\gamma}) \geq 0$, which is equivalent to the inequality in Eq. 17. □

Proposition 4 (Weak budget balance)**.** *FVCG is weakly budget balanced (WBB) i.f.f. the usefulness of the original datasets and the true cost types satisfy*

$$\begin{aligned}
&\sum_{i=0}^{n-1} [h_i(\boldsymbol{d}_{-i}, \boldsymbol{\gamma}_{-i}) + g_i(d_i)] \\
&\qquad \leq S^*(\boldsymbol{d}, \boldsymbol{\gamma}) - \sum_{i=0}^{n-1} [S^*(\boldsymbol{d}, \boldsymbol{\gamma}) - S_{-i}^*(\boldsymbol{d}_{-i}, \boldsymbol{\gamma}_{-i})]. \qquad (19)
\end{aligned}$$

Proof. Since playing truthfully is shown to be the dominant strategy by Proposition 1, we use $\boldsymbol{d}, \boldsymbol{\gamma}$ to substitute $\check{\boldsymbol{d}}, \hat{\boldsymbol{\gamma}}$ in Eq. 2 and Eq. 4. Then, the total payment to all participants is

$$\begin{aligned}
\sum_{i=0}^{n-1} p_i(\boldsymbol{d}, \boldsymbol{\gamma}) &= \sum_{i=0}^{n-1} [\tau_i(\boldsymbol{d}, \boldsymbol{\gamma}) + h_i(\boldsymbol{d}_{-i}, \boldsymbol{\gamma}_{-i}) + g_i(d_i)] \\
&= \sum_{i=0}^{n-1} [S^*(\boldsymbol{d}, \boldsymbol{\gamma}) - S_{-i}^*(\boldsymbol{d}_{-i}, \boldsymbol{\gamma}_{-i}) + c_i(d_i \eta_i^*(\boldsymbol{d}, \boldsymbol{\gamma}), \gamma_i) \\
&+ h_i(\boldsymbol{d}_{-i}, \boldsymbol{\gamma}_{-i}) + g_i(d_i)]. \qquad (20)
\end{aligned}$$

WBB means $\sum_{i=0}^{n-1} p_i(\boldsymbol{d}, \boldsymbol{\gamma}) \leq B(\boldsymbol{d} \odot \boldsymbol{\eta}^*(\boldsymbol{d}, \boldsymbol{\gamma}))$. This inequality, together with the definition $S^*(\boldsymbol{d}, \boldsymbol{\gamma}) = B(\boldsymbol{d} \odot \boldsymbol{\eta}^*(\boldsymbol{d}, \boldsymbol{\gamma})) - \sum_{i=0}^{n-1} c_i(d_i \odot \eta_i^*(\boldsymbol{d}, \boldsymbol{\gamma}), \gamma_i)$, transforms Eq. 20 to Eq. 19. □

According to Proposition 3 and 4, in order to find $\boldsymbol{h}^*(\cdot)$ and $\boldsymbol{g}^*(\cdot)$ that best satisfy IR and WBB, we minimize the expected difference between the left side and the right side of 17 and Eq. 19 when they are violated , i.e., we set Loss1 and Loss2 as follows:

$$\begin{aligned}
\text{Loss1} = \sum_{i=0}^{n-1} &\text{ReLu}[-(S^*(\boldsymbol{d}, \boldsymbol{\gamma}) - S_{-i}^*(\boldsymbol{d}_{-i}, \boldsymbol{\gamma}_{-i})) \\
&- (h_i(\boldsymbol{d}_{-i}, \boldsymbol{\gamma}_{-i}) + g_i(d_i))], \qquad (21) \\
\text{and} \quad \text{Loss2} = &\text{ReLu}[\sum_{i=0}^{n-1} [(S^*(\boldsymbol{d}, \boldsymbol{\gamma}) - S_{-i}^*(\boldsymbol{d}_{-i}, \boldsymbol{\gamma}_{-i})) \\
&+ (h_i(\boldsymbol{d}_{-i}, \boldsymbol{\gamma}_{-i}) + g_i(d_i))] - S^*(\boldsymbol{d}, \boldsymbol{\gamma})]. \qquad (22)
\end{aligned}$$

Substituting Eq. 21 and 22 into Eq. 5, we get the objective function of the composite network in Eq. 7.

Depending on different assumptions, Loss1 and Loss2 may or may not be reduced to zero by learning the optimal $\boldsymbol{h}^*(\cdot)$ and $\boldsymbol{g}^*(\cdot)$. Nevertheless, for those $\boldsymbol{h}^*(\cdot)$ and $\boldsymbol{g}^*(\cdot)$ that make Eq. 17 and 19 hold together for arbitrary $(\boldsymbol{d}, \boldsymbol{\gamma})$, we have Loss1 $\equiv$ Loss2 $\equiv 0$, and hence IR and WBB are satisfied

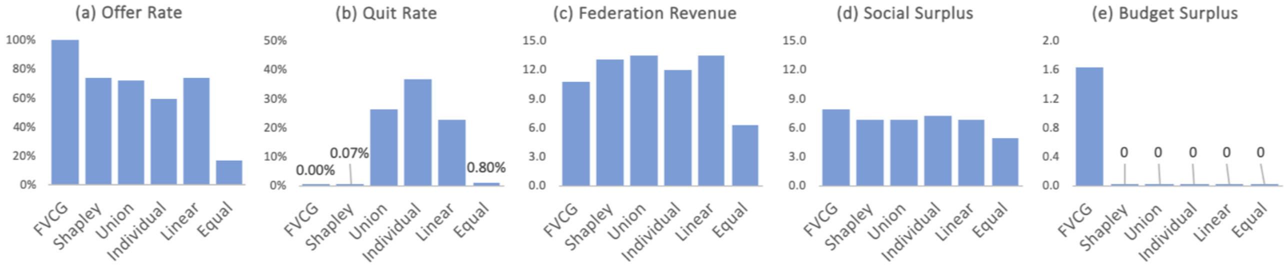

Figure 3: Comparison between different FL payoff schemes

universally. We can prove that the following inequality is a sufficient condition for the existence of such $\boldsymbol{h}^*(\cdot), \boldsymbol{g}^*(\cdot)$:

$$\sum_{i=0}^{n-1}[S^*(\boldsymbol{d},\boldsymbol{\gamma}) - S^*_{-i}(\boldsymbol{d}_{-i},\boldsymbol{\gamma}_{-i})] \leq S^*(\boldsymbol{d},\boldsymbol{\gamma}), \quad \forall \boldsymbol{d},\boldsymbol{\gamma}, \tag{23}$$

where $S^*_{-i}(\boldsymbol{d}_{-i},\boldsymbol{\gamma}_{-i})$ is the maximum social surplus attained by the $n-1$ participants except for i.

The following two assumptions lead to Eq. 23, thus providing guarantees for IR and WBB: 1) the federation revenue function is super additive (i.e., $B(\hat{\boldsymbol{d}}) \geq \sum_{i=0}^{n-1} B(\hat{d}_i), \forall \hat{\boldsymbol{d}}$) and 2) the marginal effect of the usefulness of one participant's data decreases as the usefulness of other participants' data increases (i.e., $B(\hat{d}_i, \hat{\boldsymbol{d}}_{-i}) - B(\hat{d}'_i, \hat{\boldsymbol{d}}_{-i}) \leq B(\hat{d}_i, \hat{\boldsymbol{d}}'_{-i}) - B(\hat{d}'_i, \hat{\boldsymbol{d}}'_{-i}), \forall \hat{d}_i \geq \hat{d}'_i, \hat{\boldsymbol{d}}_{-i} \geq \hat{\boldsymbol{d}}'_{-i}$). These two assumptions apply to most FL application scenarios. Proofs of Eq. 23 and its relationship with the aforementioned assumptions are beyond the scope of this paper and hence are omitted here.

6 Experimental Evaluation

In this section, we experimentally evaluate the performance of FVCG against five existing payoff schemes. They are:

1. *Shapley*: the federation revenue is shared among participants according to the Shapley value [Jia *et al.*, 2019];
2. *Union*: participant i's share of the federation revenue follows the Labour Union game [Gollapudi *et al.*, 2017] payoff scheme and is proportional to its marginal contribution to the revenue of the federation formed by its predecessors, i.e., $B((d_0, \ldots, d_i)) - B((d_0, \ldots, d_{i-1}))$;
3. *Individual*: participant i's share of the federation revenue is proportional to its marginal contribution to the federation revenue [Yang *et al.*, 2017];
4. *Linear*: participant i's share of the federation revenue is proportional to the usefulness of its contributed data (this is a payoff scheme that we designed for experimental comparison purposes only); and
5. *Equal*: the federation revenue is equally divided among participants in this federation [Yang *et al.*, 2017].

We built a simulator which creates participants with diverse characteristics to study how the different payoff schemes perform. The relative performance of the payoff schemes in the simulation is more important than the exact values.

6.1 Experiment Settings

In the experiment, we carry out simulations based on hypothetical revenue and cost functions as follows:

$$B(\hat{\boldsymbol{d}}) = \sqrt{n(\sum_{i=0}^{n-1}\hat{d}_i)}, \text{ and } c_i(\hat{d}_i, \gamma_i) = \gamma_i \hat{d}_i, i \in N. \tag{24}$$

The user-defined unfairness function is set to be that in Eq. 6. We set $n = 10$ and $\Delta(d_i), \Delta(\gamma_i)$, the prior belief on usefulness and cost types, to be one-dimensional uniform distributions on $[0, 5]$ and $[0, 1]$, respectively. We set the the hyper-parameters in Algorithm 1 to be $\lambda_1 = 0.3, \lambda_2 = 0.3, \lambda_3 = 0.4, T = 100, a = 0.001, b = 1.0$, and NET^h_i, NET^g_i to have three 10-dimensional hidden layers and one 50-dimensional hidden layer, respectively.

We run simulations to compare the FVCG mechanism with other payoff schemes. We simulate 1,000 samples for 10 participants. In each sample t, we randomly generate the usefulness d^t_i of the original dataset and the true cost type γ^t_i from $\Delta(d_i), \Delta(\gamma_i)$. For each payoff scheme, we assume at the beginning, each participant contributes (or offers) $k = 10\%$ of its data. In each iteration s, based on other participants' behaviors, each participant choose between 1) to stop offering/contributing data; 2) to offer/contribute another 10% of its data; 3) to withdraw 10% of its data from the federation; or 4) to maintain the status quo. The iteration stops when equilibrium is reached or after 1,000 iterations. For FVCG, we further assume each participant randomly reports a false cost type at first and adjust the reported cost type in each iteration.

After the simulations reach the final state, we calculate the average of the following measures across all participants and all samples:

1. *Offer Rate*: the ratio between the usefulness of the *offered* dataset to that of the original dataset, $\check{d}^t_i / d^t_i$ (for payoff schemes other than FVCG, $\check{d}^t_i = \hat{d}^t_i$);
2. *Quit Rate*: (No. of participants that quit the federation) / (No. of total participants);
3. *Federation Revenue*: $B(\hat{\boldsymbol{d}}^t)$;
4. *Social Surplus*: $S(\hat{\boldsymbol{d}}^t, \boldsymbol{\gamma}^t) = B(\hat{\boldsymbol{d}}^t) - \sum_{i=0}^{n-1} c_i(\hat{d}_i, \gamma_i)$, which is the total utilities of all participants (including the FL coordinator); and
5. *Budget Surplus*: $B(\hat{\boldsymbol{d}}^t) - \sum_{i=0}^{n-1} p^t_i$, which is also the profit gained by the FL coordinator.

6.2 Results and Discussion

Experimental evaluation results (Figure 3) are consistent with our theoretical predictions. The FVCG mechanism incentivizes participants to offer all their data (i.e., the offer rate is 100%) and keeps all participants in the federation (i.e., the quit rate is 0). Although the federation revenue of FVCG is lower than other payoff schemes (because FVCG rejects data if the marginal contribution cannot cover the marginal cost), FVCG results in the highest social surplus and significant budget surplus (around 15% of the federation revenue). This means the federation as a whole has the highest excess economic gain over costs, and the FL coordinator can make considerable profits by adjusting FL participation.

7 Conclusions and Future Work

In this paper, we presented the FVCG mechanism for federated learning. It incentivizes participants to join FL model training with their original dataset and truthfully report their cost types. It maximizes social surplus, minimizes expected deviations from individual rationality, is by expectation weakly budget balanced, and supports user-defined fairness criteria. With the theoretical guarantee of incentive compatibility, FVCG achieves a desirable trade-off among all these considerations.

In subsequent research, we will further study how to evaluate the usefulness of datasets, identify the federation revenue function, and estimate cost types in a privacy-preserving manner. Each of them is a promising research direction.

References

[Cai *et al.*, 2015] Yang Cai, Constantinos Daskalakis, and Christos Papadimitriou. Optimum statistical estimation with strategic data sources. In *Conference on Learning Theory*, pages 280–296, 2015.

[Caragiannis *et al.*, 2016] Ioannis Caragiannis, Ariel Procaccia, and Nisarg Shah. Truthful univariate estimators. In *International Conference on Machine Learning*, pages 127–135, 2016.

[Clarke, 1971] Edward H Clarke. Multipart pricing of public goods. *Public Choice*, 11(1):17–33, 1971.

[Cummings *et al.*, 2015] Rachel Cummings, Stratis Ioannidis, and Katrina Ligett. Truthful linear regression. In *Conference on Learning Theory*, pages 448–483, 2015.

[Dekel *et al.*, 2010] Ofer Dekel, Felix Fischer, and Ariel D Procaccia. Incentive compatible regression learning. *Journal of Computer and System Sciences*, 76(8):759–777, 2010.

[Faltings and Radanovic, 2017] Boi Faltings and Goran Radanovic. Game theory for data science: eliciting truthful information. *Synthesis Lectures on Artificial Intelligence and Machine Learning*, 11(2):1–151, 2017.

[Fleischer and Lyu, 2012] Lisa K Fleischer and Yu-Han Lyu. Approximately optimal auctions for selling privacy when costs are correlated with data. In *Proceedings of the 13th ACM Conference on Electronic Commerce*, pages 568–585. ACM, 2012.

[Ghosh and Roth, 2015] Arpita Ghosh and Aaron Roth. Selling privacy at auction. *Games and Economic Behavior*, 91:334–346, 2015.

[Gollapudi *et al.*, 2017] Sreenivas Gollapudi, Kostas Kollias, Debmalya Panigrahi, and Venetia Pliatsika. Profit sharing and efficiency in utility games. In *Proceedings of the 25th Annual European Symposium on Algorithms (ESA'17)*, 2017.

[Groves, 1973] Theodore Groves. Incentives in teams. *Econometrica*, 41(4):617–631, 1973.

[Jackson, 2014] Matthew O Jackson. Mechanism theory. *Available at SSRN 2542983*, 2014.

[Jia *et al.*, 2019] Ruoxi Jia, David Dao, Boxin Wang, Frances Ann Hubis, Nick Hynes, Nezihe Merve Gurel, Bo Li, Ce Zhang, Dawn Song, and Costas Spanos. Towards efficient data valuation based on the shapley value. *CoRR, arXiv:1902.10275*, 2019.

[Kairouz *et al.*, 2019] Peter Kairouz, H Brendan McMahan, Brendan Avent, et al. Advances and open problems in federated learning. In *CoRR*, page arXiv:1912.04977, 2019.

[Koh and Liang, 2017] Pang Wei Koh and Percy Liang. Understanding black-box predictions via influence functions. In *Proceedings of the 34th International Conference on Machine Learning-Volume 70*, pages 1885–1894. JMLR. org, 2017.

[Koutsoupias, 2014] Elias Koutsoupias. Scheduling without payments. *Theory of Computing Systems*, 54(3):375–387, 2014.

[Meir *et al.*, 2012] Reshef Meir, Ariel D Procaccia, and Jeffrey S Rosenschein. Algorithms for strategyproof classification. *Artificial Intelligence*, 186:123–156, 2012.

[Nisan and Ronen, 2007] Noam Nisan and Amir Ronen. Computationally feasible vcg mechanisms. *Journal of Artificial Intelligence Research*, 29:19–47, 2007.

[Nisan *et al.*, 2007] Noam Nisan, Tim Roughgarden, Eva Tardos, and Vijay V Vazirani. *Algorithmic game theory*. Cambridge university press, 2007.

[Nissim *et al.*, 2012] Kobbi Nissim, Claudio Orlandi, and Rann Smorodinsky. Privacy-aware mechanism design. In *Proceedings of the 13th ACM Conference on Electronic Commerce*, pages 774–789. ACM, 2012.

[Perote and Perote-Pena, 2004] Javier Perote and Juan Perote-Pena. Strategy-proof estimators for simple regression. *Mathematical Social Sciences*, 47(2):153–176, 2004.

[Richardson *et al.*, 2019] Adam Richardson, Aris Filos-Ratsikas, and Boi Faltings. Rewarding high-quality data via influence functions. *CoRR, arXiv:1908.11598*, 2019.

[Vickrey, 1961] William Vickrey. Counterspeculation, auctions, and competitive sealed tenders. *The Journal of finance*, 16(1):8–37, 1961.

[Wang, 2019] Guan Wang. Interpret federated learning with shapley values. *the 1st International Workshop on Federated Machine Learning for User Privacy and Data Confidentiality*, 2019.

[Westenbroek *et al.*, 2019] Tyler Westenbroek, Roy Dong, Lillian J Ratliff, and S Shankar Sastry. Competitive statistical estimation with strategic data sources. *CoRR, arXiv:1904.12768*, 2019.

[Yang *et al.*, 2017] Shuo Yang, Fan Wu, Shaojie Tang, Xiaofeng Gao, Bo Yang, and Guihai Chen. On designing data quality-aware truth estimation and surplus sharing method for mobile crowdsensing. *IEEE Journal on Selected Areas in Communications*, 35(4):832–847, 2017.

[Yang *et al.*, 2019a] Qiang Yang, Yang Liu, Tianjian Chen, and Yongxin Tong. Federated machine learning: Concept and applications. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 10(2):12, 2019.

[Yang *et al.*, 2019b] Qiang Yang, Yang Liu, Yong Cheng, Yan Kang, Tianjian Chen, and Han Yu. *Federated Learning*. Morgan & Claypool Publishers, 2019.

[Yu *et al.*, 2020] Han Yu, Zelei Liu, Yang Liu, Tianjian Chen, Mingshu Cong, Dusit Niyato, and Yang Qiang. A fairness-aware incentive scheme for federated learning. In *Proceedings of the 3rd AAAI/ACM Conference on Artificial Intelligence, Ethics, and Society (AIES-20)*, 2020.